DEPARTMENT OF THE ARMY
U.S. Army Edgewood Chemical Biological Center
Aberdeen Proving Ground, Maryland 21010-5424

ERRATUM SHEET

06 Oct 2010

| | |
|---|---|
| REPORT NO. | ECBC-TR-771 |
| TITLE | Discrimination of Pathogenic Versus Non-pathogenic *Yersinia pestis* and *Escherichia coli* using Proteomics Mass Spectrometry |
| AUTHORS | Mary Margaret Wade, Alan W. Zulich, Charles H. Wick (ECBC), Rabih E. Jabbour (SAIC), Samir V. Deshpande (STC) |
| DATE | May 2010 |
| CLASSIFICATION | UNCLASSIFIED - Unlimited |

In copies of ECBC-TR-771 previously sent to you, please replace the cover and inside cover, and pages 1 and 2 with the attached replacement pages.

Sandra J. Johnson
SANDRA J. JOHNSON
Chief, Technical Releases Office

Attachments

20100624420

A522639


# EDGEWOOD

## CHEMICAL BIOLOGICAL CENTER

U.S. ARMY RESEARCH, DEVELOPMENT AND ENGINEERING COMMAND

**ECBC-TR-771**

# DISCRIMINATION OF PATHOGENIC VERSUS NON-PATHOGENIC *YERSINIA PESTIS* AND *ESCHERICHIA COLI* USING PROTEOMICS MASS SPECTROMETRY

**Mary Margaret Wade**
**Alan W. Zulich**
**Charles H. Wick**

**RESEARCH AND TECHNOLOGY DIRECTORATE**


**Rabih E. Jabbour**

**SCIENCE APPLICATIONS INTERNATIONAL CORPORATION**
**Gunpowder, MD 21010-0068**

**Samir V. Deshpande**


**SCIENCE AND TECHNOLOGY CORPORATION**
**Edgewood, MD 21040**

**May 2010**



**ABERDEEN PROVING GROUND, MD 21010-5424**

## Disclaimer

The findings in this report are not to be construed as an official Department of the Army position unless so designated by other authorizing documents.

# REPORT DOCUMENTATION PAGE

*Form Approved*
*OMB No. 0704-0188*



| 1. REPORT DATE *(DD-MM-YYYY)* | 2. REPORT TYPE | 3. DATES COVERED *(From - To)* |
|---|---|---|
| XX-05-2010 | Final | Jan 2009 - Sep 2009 |

**4. TITLE AND SUBTITLE**

Discrimination of Pathogenic Versus Non-pathogenic *Yersinia pestis* and *Escherichia coli* Using Proteomics Mass Spectrometry

**5a. CONTRACT NUMBER**

**5b. GRANT NUMBER**

**5c. PROGRAM ELEMENT NUMBER**

**6. AUTHOR(S)**
Wade, Mary Margaret; Zulich, Alan W.; Wick, Charles H. (ECBC); Jabbour, Rabih E. (SAIC); and Deshpande, Samir V. (STC).

**5d. PROJECT NUMBER**
None

**5e. TASK NUMBER**

**5f. WORK UNIT NUMBER**

**7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)**
DIR, ECBC, ATTN: RDCB-DRB-D, APG, MD 21010-5424
SAIC, P.O. Box 68, Gunpowder, MD 21010-0068
STC, 500 Edgewood Road, Suite 205, Edgewood, MD 21040

**8. PERFORMING ORGANIZATION REPORT NUMBER**
ECBC-TR-771

**9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)**

**10. SPONSOR/MONITOR'S ACRONYM(S)**

**11. SPONSOR/MONITOR'S REPORT NUMBER(S)**

**12. DISTRIBUTION / AVAILABILITY STATEMENT**
Approved for public release; distribution is unlimited.

**13. SUPPLEMENTARY NOTES**

**14. ABSTRACT**

Recently, the use of mass spectrometry analysis to characterize and identify biological agents using a proteomic approach has been demonstrated. The present study sought to determine whether proteomics mass spectrometry could be used to distinguish between pathogenic and non-pathogenic strains of the same organism based on outer membrane proteins (OMPs). OMPs are often associated with virulence in gram-negative pathogens and could prove to be excellent model biomarkers for strain differentiation among bacteria. OMPs were isolated from pathogenic and non-pathogenic strains of *Yersinia pestis* and *Escherichia coli* using ultracentrifugation and a sarkosyl extraction method followed by protein digestion and analysis using liquid chromatography-mass spectrometry/mass spectrometry (LC-MS/MS). The experimental MS/MS spectral data of bacterial peptides were searched using the SEQUEST algorithm against a constructed proteome database of microorganisms to determine the identity and number of unique peptides. Results showed that pathogenic strains could be distinguished from non-pathogenic strains for both organisms based on differences in OMP profiles. Virulent *E. coli* and *Y. pestis* expressed multiple OMPs that were not present in avirulent samples. In conclusion, proteomics MS proved useful in discriminating between pathogenic and non-pathogenic strains, and OMPs unique to pathogenic strains could prove to be useful biomarkers for strain differentiation.

**15. SUBJECT TERMS**
Yersinia pestis; Escherichia coli; Outer membrane proteins; Proteomics; Mass spectrometry

| 16. SECURITY CLASSIFICATION OF: | | | 17. LIMITATION OF ABSTRACT | 18. NUMBER OF PAGES | 19a. NAME OF RESPONSIBLE PERSON |
|---|---|---|---|---|---|
| a. REPORT | b. ABSTRACT | c. THIS PAGE | | | Sandra J. Johnson |
| U | U | U | UL | 22 | 19b. TELEPHONE NUMBER *(include area code)* (410) 436-2914 |

Blank

## PREFACE

The work described in this report was started in January 2009 and completed in September 2009.

The use of either trade or manufacturers' names in this report does not constitute an official endorsement of any commercial products. This report may not be cited for purposes of advertisement.

This report has been approved for public release. Registered users should request additional copies from the Defense Technical Information Center; unregistered users should direct such requests to the National Technical Information Service.


### Acknowledgment

This work was funded by the 2009 U.S. Army Edgewood Chemical Biological Center's In-house Laboratory Independent Research program.

Blank

CONTENTS


1. INTRODUCTION ..... 7

2. MATERIALS AND METHODS ..... 8

2.1 Materials and Reagents ..... 8
2.2 Bacterial Strains and Culture Conditions ..... 8
2.3 OMP Isolation ..... 8
2.4 Processing of Whole Cell Lysates and OMP Samples ..... 9
2.5 Protein Database and Database Search Engine ..... 9

3. RESULTS AND DISCUSSION ..... 10

3.1 Differentiation of Pathogenic vs. Non-pathogenic *E. coli* Strains Using Whole Cell Lysates ..... 12
3.2 Differentiation of Pathogenic vs. Non-pathogenic *E. coli* Strains Using OMPs ..... 13
3.3 Differentiation of Pathogenic vs. Non-pathogenic *Y. pestis* Strains Using Whole Cell Lysates and OMPs ..... 15
3.4 Differentiation of Pathogenic vs. Non-pathogenic *B. anthracis* Strains Using Whole Cell Lysates ..... 18

4. CONCLUSIONS ..... 20

LITERATURE CITED ..... 21

## FIGURES


1. MS-based Proteomic Approach Output ..... 11

2. Histogram Representing the Output of the Binary Matrix of the Unique Peptides Identified for a Given Bacterium at 95% Confidence Level ..... 11

3. Near-neighbor Classification of Pathogenic *E. coli* O157:H7 vs. Non-pathogenic *E. coli* K12 Using Whole Cell Lysates ..... 12

4. Near-neighbor Classification of Non-pathogenic *E. coli* K12 vs. Pathogenic *E. coli* O157:H7 Using OMP Extracts ..... 15

5. Near-neighbor Classification of Non-pathogenic *Y. pestis* A1122 vs. Pathogenic *Y. pestis* Using OMP Extracts ..... 17

6. Bacterial Differentiation of *B. anthracis* Strains Using Whole Cell Lysates ..... 19


## TABLES


1. Unique Proteins in *E. coli* O157:H7 vs. *E. coli* K12 (Potential Biomarkers) ..... 14

2. Unique Proteins for *Y. pestis* from Whole Cell Lysates vs. OMP Extracts ..... 18

# DISCRIMINATION OF PATHOGENIC VERSUS NON-PATHOGENIC *YERSINIA PESTIS* AND *ESCHERICHIA COLI* USING PROTEOMICS MASS SPECTROMETRY

## 1. INTRODUCTION

Recently, mass spectrometry (MS) analysis has proven useful in the characterization and identification of biological agents using a proteomic approach (1). Therefore, the present study sought to determine whether proteomics MS could be used to distinguish between pathogenic and non-pathogenic strains of the same organism. More specifically, discrimination between pathogenic and non-pathogenic organisms based on their outer membrane protein (OMP) composition, as determined by MS, was investigated.

OMPs of gram-negative bacteria act as active mediators between the cell and its environment and are often associated with virulence in gram-negative pathogens. In pathogenic *Escherichia coli*, there are multiple OMPs present, which are needed for intestinal colonization, as well as those that play a role in the type III secretion system responsible for delivering effector proteins to host cells (2, 3, 4, 5, 6). Virulent *Yersinia pestis* contains three plasmids encoding multiple OMPs that are required for virulence (7, 8, 9). For example the pCD1 plasmid encodes several *Yersinia* outer membrane proteins (YOPs) and a type III secretion system, which are needed for survival and entry into eukaryotic cells (10, 11). Additionally, the pPCP1 plasmid encodes an OMP plasminogen activator that interferes with clotting and complement (12). Avirulent strains often lack one or more of the plasmids or genes encoding proteins needed for virulence, and it is these differences in OMP expression between virulent and avirulent strains of gram-negative bacteria that could potentially be exploited to distinguish among strains. Therefore, OMPs could prove to be excellent model biomarkers for strain differentiation among bacteria.

The objective of the present study was to establish the sequence-based identity of OMPs isolated from pathogenic and non-pathogenic strains of *Y. pestis* and *E. coli*. *Y. pestis* is classified as a Category A pathogen and is an important potential biological warfare agent. Pathogenic *E. coli*, such as *E. coli* O157:H7, is an important public health pathogen responsible for most common food borne and waterborne illnesses in the United States. High-throughput proteomic analytical systems were applied, providing a rapid means of characterizing cellular proteins and producing amino acid sequence information for peptides derived from these proteins.

This 1 year basic research study aimed to: 1) isolate OMPs using ultracentrifugation and differential extractions; 2) determine sequence and post-translational modifications to amino acid residues composing membrane proteins using emerging high-throughput mass spectral proteomic systems; and 3) use bioinformatics modeling tools to establish strain differentiation methods based on the proteome differences among the *Y. pestis* and *E. coli* strains.

In addition to the aims described above, discrimination among strains of an additional agent of interest, *Bacillus anthracis*, was also investigated. Because *B. anthracis* is a gram-positive organism and therefore, lacks an outer membrane, total cellular proteins (whole cell lysates) were analyzed for discrimination via mass spectrometry rather than OMPs.

## 2. MATERIALS AND METHODS

### 2.1 Materials and Reagents.

Ammonium bicarbonate, dithiotheritol, urea, acetonitrile-HPLC grade, and formic acid were purchased from Burdick and Jackson (St. Louis, MO). Sequencing grade modified trypsin was purchased from Promega (Madison, WI).

### 2.2 Bacterial Strains and Culture Conditions.

Pathogenic strains used in the present study were: *E. coli* O157:H7, *Y. pestis* Colorado 92 (CO92), and *B. anthracis* Ames. Non-pathogenic strains used were *E. coli* K12, *Y. pestis* A1122, and *B. anthracis* Sterne. Working cultures were prepared by streaking cells from cryo-preserved stocks onto tryptic soy agar (TSA) followed by incubation for approximately 18 h at 37 °C for *E. coli* and *B. anthracis* strains and 30 °C for *Y. pestis* strains. After incubation, all working culture plates were stored at 4 °C. Cells from working cultures were used to inoculate broth cultures for each strain, which consisted of 100 mL of tryptic soy broth (TSB) for *E. coli* and *B. anthracis* strains and 100 mL of brain heart infusion (BHI) for *Y. pestis* strains. All cultures were incubated for approximately 18 h at 37 °C for *E. coli* and *B. anthracis* strains and 30 °C for *Y. pestis* strains with rotary aeration at 180 rpm. After incubation, broth cultures were pelleted by centrifugation (2,300 RCF at 4 °C for 10 min), washed and resuspended in 10 mL HEPES buffer followed by heating at 95 °C for 1 h to lyse cells. After heating, a portion of each sample was plated onto TSA and incubated for 5 days at appropriate temperature to ensure no growth prior to removing samples from the BSL-2 or BSL-3 laboratory for further processing. Total cellular protein samples (whole cell lysates) were complete after heating for the 1 h and were transferred to Point Detection Branch for analysis after no growth on plates was confirmed. For OMP samples, samples were processed for OMP isolation as described below prior to being transferred to Point Detection Branch for analysis.

### 2.3 OMP Isolation.

After lysis by heating at 95 °C for 1 h, cell debris was pelleted by centrifugation at 2,300 RCF at 4 °C for 10 min. The supernatant was then centrifuged at 100,000 x g for 1 h to pellet proteins. The pellet was resuspended in 1 mL of HEPES buffer, and 1 mL of a 2% sarkosyl solution (N-Lauroylsarcosine sodium salt solution) was added and sample was incubated at room temperature for 30 min. Next samples were centrifuged at 100,000 x g for 1 h, and the pellet containing OMPs was resuspended in 1 mL of HEPES buffer and then transferred to Point Detection Branch for further processing and analysis as described below.

## 2.4 Processing of Whole Cell Lysates and OMP Samples.

All protein samples were ultra-sonicated (20 s pulse on, 5 s pulse off, and 25% amplitude for 5 min duration), and a small portion of lysates was reserved for 1-D gel analysis. The lysates were centrifuged at 14,100 x g for 30 min to remove any debris. The supernatant was then added to a Microcon YM-3 filter unit (Millipore; Cat #: 42404) and centrifuged at 14,100 x g for 30 min. The effluent was discarded. The filter membrane was washed with 100 mM ammonium bicarbonate (ABC) and centrifuged for 15-20 min at 14,100 x g. Proteins were denatured by adding 8 M urea and 3 μg/μL Dithiotheritol (DTT) to the filter and incubating overnight at 37 °C on an orbital shaker set to 60 rpm. Twenty microliters of 100% of acetonitrile (ACN) was added to the tubes and allowed incubate at room temperature for 5 min. The tubes were then centrifuged at 14,100 x g for 30-40 min and washed three times using 150 μL of 100 mM ABC solution. On the last wash, ABC was allowed to sit on the membrane for 20 min while shaking, followed by centrifugation at 14,100 x g for 30-40 min. The micron filter unit was then transferred to a new receptor tube and proteins were digested with 5 μL trypsin in 240 μL of ABC solution + 5 μL ACN. Proteins were digested overnight at 37 °C on an orbital shaker set to 55 rpm. Sixty microliters of 5% ACN/0.5% formic acid (FA) was added to each filter to quench the trypsin digestion followed by 2 min of vortexing for sample mixing. The tubes were centrifuged for 20-30 min at 14,100 x g. An additional 60 μL 5% ACN/0.5% FA mixture was added to filter and centrifuged. The effluent was then analyzed using the LC-MS/MS technique.

## 2.5 Protein Database and Database Search Engine.

A protein database was constructed in a FASTA format using the annotated bacterial proteome sequences derived from fully sequenced chromosomes of 881 bacteria, including their sequenced plasmids (as of April 2009). A PERL program (http://www.activestate.com/Products/ActivePerl; accessed April 2009) was written to automatically download these sequences from the National Institutes of Health National Center for Biotechnology (NCBI) site (http://www.ncbi.nlm.nih.gov; accessed April 2009). Each database protein sequence was supplemented with information about a source organism and a genomic position of the respective ORF embedded into a header line. The database of bacterial proteomes was constructed by translating putative protein-coding genes and consists of tens of millions of amino acid sequences of potential tryptic peptides obtained by the *in silico* digestion of all proteins (assuming up to two missed cleavages).

The experimental MS/MS spectral data of bacterial peptides were searched using SEQUEST algorithm against a constructed proteome database of microorganisms. The SEQUEST thresholds for searching the product ion mass spectra of peptides were Xcorr, deltaCn, Sp, RSp, and deltaMpep. These parameters provided a uniform matching score of all candidate peptides. The generated outfiles of these candidate peptides were then validated using a peptide prophet algorithm. Peptide sequences with a probability score of 95% and higher were retained in the dataset and used to generate a binary matrix of sequence-to-bacterium assignments. The binary matrix assignment was populated by matching the peptides with corresponding proteins in the database and assigning a score of 1. A score of zero was assigned for a non match. The column in the binary matrix represents the proteome of a given bacterium

and each row represents a tryptie peptide sequenee from the LC-MS/MS analysis. Mieroorganisms were matehed with the baeterium/baeteria based on the number of unique peptides that remained after further filtering of degenerate peptides from the binary matrix. Verifieation of the elassifieation and identifieation of eandidate mieroorganisms was performed through hierarehieal elustering analysis and taxonomie elassifieation.

The in-house developed software ealled "BACid" transformed results of searehing MS/MS speetra of peptide ions against a eustom protein database whieh was downloaded from NCBI with eommereial software SEQUEST into a taxonomieally meaningful and easy to interpret output. It ealculated probabilities that peptide sequenee assignment to a MS/MS speetrum was eorreet and used aeeepted speetrum–to-sequenee matehes to generate a sequenee-to-baeterium (STB) binary matrix of assignments. Validated peptide sequenees, differentially present or absent in various strains (STB matriees) were visualized as assignment bitmaps and analyzed by a BACid module that used phylogenetie relationships among baeterial speeies as a part of deeision tree proeess. The baeterial elassifieation and identifieation algorithm used assignments of organisms to taxonomie groups (phylogenetie elassifieation) based on an organized seheme that begins at the phylum level and follows through elasses, orders, families and genus down to strain level. BACid was developed in-house using PERL, MATLAB and Mierosoft Visual Basie.

## 3. RESULTS AND DISCUSSION

The eurrent projeet eharaeterized and identified pathogenie and non-pathogenie strains of the same organism based on proteins present in whole eell lysates (global) versus OMP preparations (speeifie). All results are shown and diseussed below. *B. anthracis* Ames and Sterne strains were also ineluded to expand the projeet; however, analysis of whole eell lysates only were ineluded and results diseussed below.

Figure 1 below serves as an example to illustrate the typieal output generated for the LC-ESI MS/MS analyses of baeterial proteins digest using bioinformaties tools to proeess the peptide sequenee information for the baeterial differentiation and elassifieation. The top window lists the identified unique proteins and their eorresponding baeterium mateh. The lower window represents the binary matrix of the sequenee-to-baeterium seareh matehing. The total row, lower window, represents the total number of unique proteins identified for a given baeterium. Figure 2 also serves as an example and shows the histogram generated by plotting the number of unique proteins versus the baeterium matehing in the database. The Y-axis represents the pereentage of unique peptides matehed with 95% eonfidenee level for all the baeteria on the x-axis. In this example ease, the identified baeterium at strain level is *Y. pestis*. The horizontal redline is the threshold eutoff under whieh eommon degenerate peptides among various baeteria within the eonstrueted proteome database are shown. These degenerate peptides are removed from the total number of unique peptides of the identified speeies.

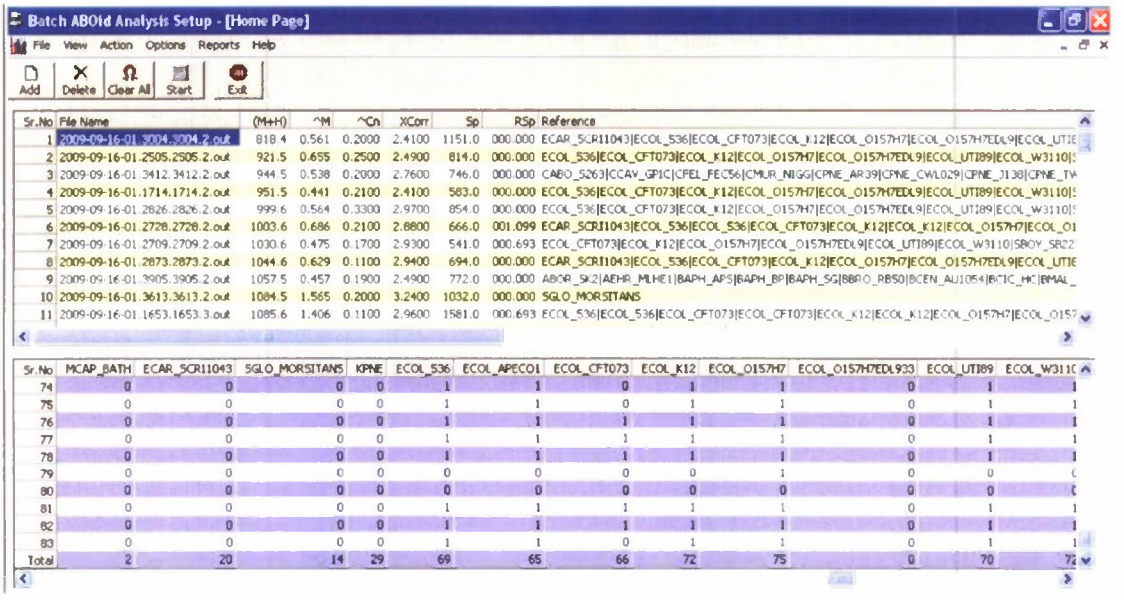


Figure 1. MS-based Proteomic Approach Output. The upper section represents the matching algorithm results of the identified tryptic peptides resulted from the LC-MS/MS analysis. Lower section represents the binary matrix of sequence-to-bacterium scoring. Presence of a unique peptide corresponding with a protein in the given proteome of a bacterium is scored 1, non match score 0.

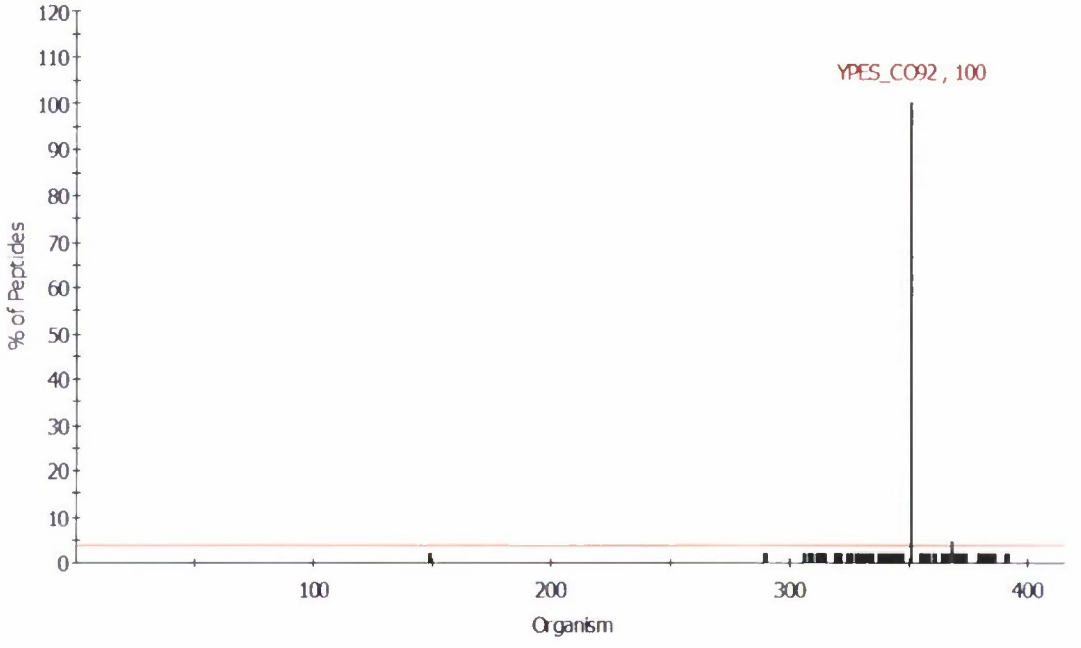


Figure 2. Histogram Representing the Output of the Binary Matrix of the Unique Peptides Identified for a Given Bacterium at 95% Confidence Level. The horizontal line is the threshold under which peptides identified are considered statistically non significant.

3.1 Differentiation of Pathogenic vs. Non-pathogenic *E. coli* Strains Using Whole Cell Lysates.

Whole cell lysates of pathogenic and non-pathogenic *E. coli* strains, *E. coli* O157:H7 and *E. coli* K12, respectively, were prepared and analyzed by proteomic mass spectrometry as described above. Results showed correct identification at the strain level for both samples analyzed. The near neighbor analysis, using Euclidean distance linkage approach, for these lysed bacterial samples showed that the identified unique set of proteins had the closest match with the employed *E. coli* strains. Therefore, correct identification to the strain level was achieved for both bacterial whole cell lysates (Figures 3 a-b).

Figure 3a shows correct identification of one sample as *E. coli* O157:H7, with the next near neighbor being *E. coli* UTI89, the causative agent of human urinary tract infections. Although *E. coli* UTI89 is closely related to *E. coli* O157:H7, it is missing certain proteins such as the BAA35715 outer membrane and flagella related proteins that are distinctly expressed in *E. coli* O157:H7, but not in *E. coli* UTI189. Moreover, the analyzed sample of the non-pathogenic *E. coli* K12, shown in Figure 3b, was correctly identified as *E. coli* K12, yet had equal similarity with *E. coli* W3110, which is a nonpathogenic strain of *E. coli* genetically very closely related to *E. coli* K12.

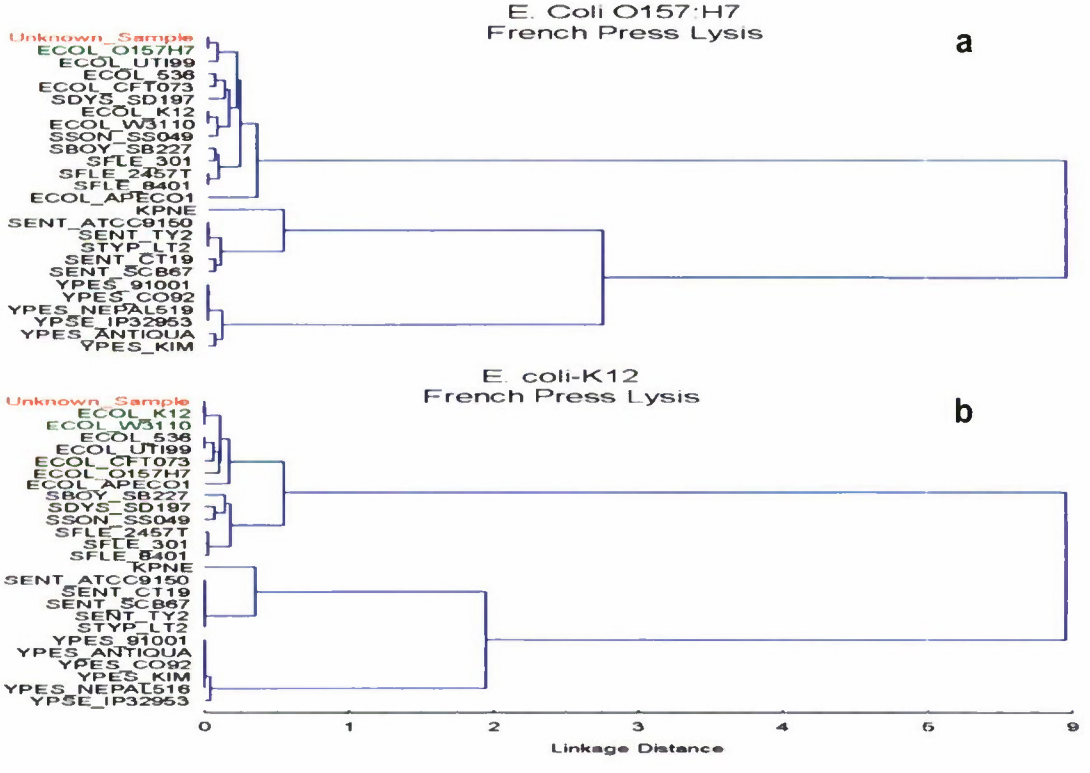


Figure 3 a-b. Near-neighbor Classification of Pathogenic *E. coli* O157:H7 (Figure 3a) vs. Non-pathogenic *E. coli* K12 (Figure 3b) Using Whole Cell Lysates.

The number of unique proteins identified differed between the pathogenic and non-pathogenic *E. coli* strains, with the pathogenic strain having a relatively lower number of unique proteins (114) than that of non-pathogeic *E. coli* (139). Among the unique proteins of *E. coli* K12, most were shared with *E. coli* O157:H7; however, the few numbers of proteins that were not shared could potentially be used for strain level discrmination. The difference in number of unique proteins could be explained by the fact that *E. coli* K12 has been extensively studied more than any other *E. coli* strain and therefore has more genetic and biochemical information available to serve as a foundation for interpreting proteome sequences from other strains (13). This difference in the number of unique proteins between the two mentioned strains probably contributed to the difference in the similarity scoring for each respective strain as shown in Table 1.

3.2 Differentiation of Pathogenic vs. Non-pathogenic *E. coli* Strains Using OMPs.

The results of using OMPs as biomarkers for bacterial differentiation of pathogenic versus non-pathogenic *E. coli* strains are shown in Figures 4a-b. Each *E. coli* strain was correctly identified with no near-neighbor strains sharing the strain level identification. Figure 4a shows the resultant near-neighbor similarity linkage analysis for OMP extracts from *E. coli* K12. The OMP extract resulted in a unique set of protein biomarkers that are capable of enhancing the differentiation at the strain level and resulted in complete similarity with *E. coli* K12 strain. No ambiguity was observed in the identification unlike that experienced when using the whole cell lysates in which an equal classification was shared between *E. coli* K12 and *E. coli* W3110 strains (Figure 3a). Although *E. coli* K12 and *E. coli* W3110 strains are genetically indistinguishable and their protein content appears very similar when analyzing whole lysates, a distinct difference was observed (significant dissimilarity) between the two closely related strains when using OMP extracts.

Figure 4b shows the near-neighbor similarity linkage results for the OMP extract of *E. coli* O157:H7. Better discrimination of *E. coli* O157:H7 was achieved using OMP extracts than that observed using whole cell lysates. The number of unique OMPs that could be identified was greater in the OMP extracts analyzed than that observed with whole cell lysates. However, this does not imply the absence of these OMPs from the whole cell lysate. Rather it is likely a higher abundance of non-OMPs in the whole cell lysate that is suppressing the detection of the OMPs in the whole lysate extracts. MS analysis has been reported to suffer ionization suppression due to the presence of large numbers of ionizable species. Generally, the whole cell lysate has a larger number of ionizable peptides and greater abundance of non-outer membrane tryptic peptides than that of OMP extracts and therefore is highly likely to experience ionization suppression during MS analysis.

Table 1. Unique Proteins in *E. coli* O157:H7 vs. *E. coli* K12 (Potential Biomarkers)

| Protein Aeeesion # | Protein Info | ***E. coli-K12 sub. HB101*** | ***E. coli-O157:H7*** |
|---|---|---|---|
| NP_418358.1 | stress-induced protein | ✓ | ✗ |
| NP_417795.1 | baeterioferritin, iron storage and detoxifieation protein | ✓ | ✗ |
| YP_671573.1 | putative eytoplasmic protein | ✓ | ✗ |
| NP_415386.1 | lipoprotein | ✓ | ✗ |
| NP_755058.1 | GnsAGnsB family protein | ✓ | ✗ |
| NP_668903.1 | Chorismate synthase | ✓ | ✗ |
| YP_670276.1 | Hypothetieal protein | ✓ | ✗ |
| YP_669714.1 | Aspartyl-tRNA synthetase | ✓ | ✗ |
| NP_312864 | two-eomponent sensor protein related to pathogcnicity islands | ✗ | ✓ |
| NP_310689.1 | Struetural flagella protein | ✗ | ✓ |
| NP_290256 | Seereted protein EspA, RELATED TO PATHOGENICITY ISLANDS | ✗ | ✓ |
| BAA35715 | Outer memebrane protein | ✗ | ✓ |
| NP_286049 | putative beta-barrel outer membrane protein | ✗ | ✓ |

Comparison of the whole cell and OMP extracts from *E. coli* O157:H7 showed distinct differences in the nature of the identified unique protein biomarkers. The whole cell lysate for *E. coli* O157:H7 had unique proteins it shared with its genetically closest strain, *E. coli* UTI89. However, this was not the case when comparing the unique proteins for these two strains using the OMP extracts. For the OMP extracts, the difference in the number of strain unique protein biomarkers increased as compared with that of the whole lysate analysis.

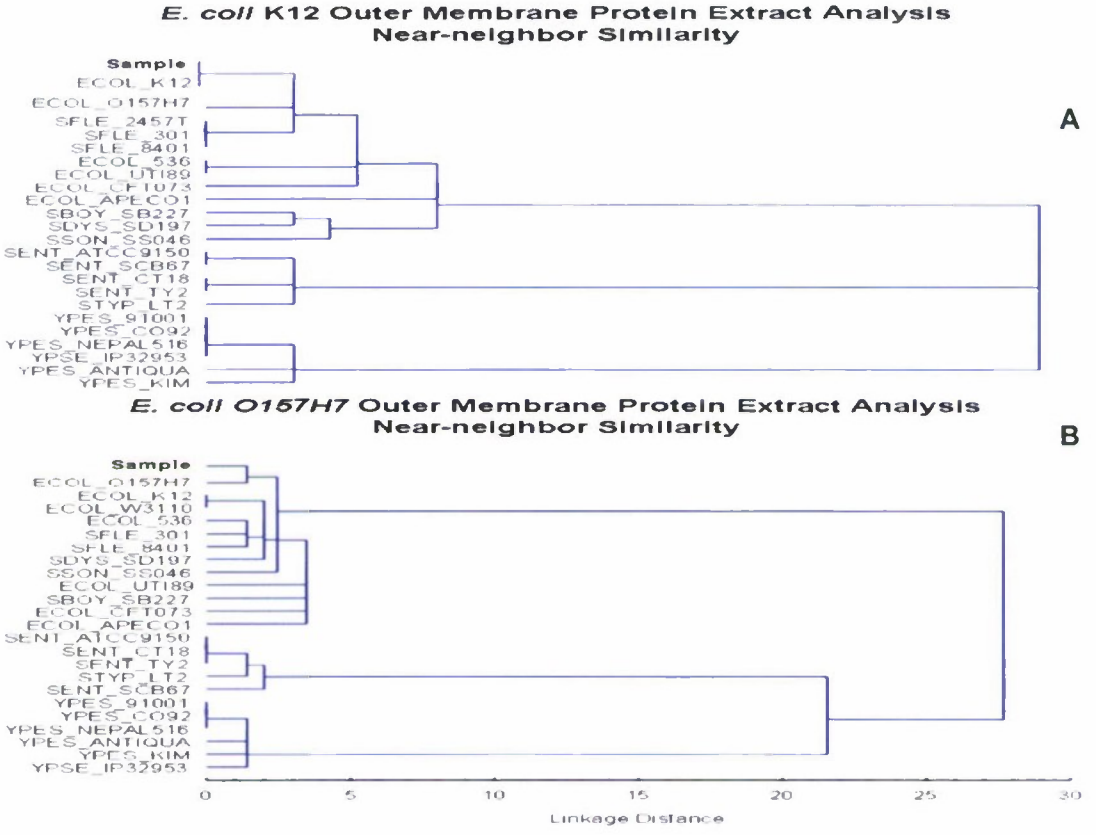


Figure 4 a-b. Near-neighbor Classification of Non-pathogenic *E. coli* K12 (Figure 4a) vs. Pathogenic *E. coli* O157:H7 (Figure 4b) Using OMP Extracts.

## 3.3 Differentiation of Pathogenic vs. Non-pathogenic *Y. pestis* Strains Using Whole Cell Lysates and OMPs.

Comparison of proteins present in whole cell lysates and OMP extracts of pathogenic versus non-pathogenic *Y. pestis* and *Y. pestis* A1122, respectively, was performed. Figure 5 a-b shows the near-neighbor analysis, using Euclidean distance linkage approach, for the bacterial identification based on OMP extracts. The identified unique sets of proteins had the closest match with the employed *Y. pestis* strains. However, it should be mentioned that the *Y. pestis* A1122 strain is not included in the current database due to the fact that its genome is not fully sequenced and publicly available. However, the constructed proteome database does

include all the pathogenic *Y. pestis* strains that are listed as pathogenic strains in the DoD classification. Figure 5a. shows the dendogram for the identification of the avirulent *Y. pestis* A1122 sample. The result showed that this strain was identified at the strain level as *Y. pestis* 91001, which is also an avirulent strain. This finding is encouraging because *Y. pestis* 91001 is the only avirulent strain in the proteome database among seven pathogenic *Y. pestis* strains currently included. Therefore, the absence of *Y. pestis* strain A1122 from the database provided an indirect test of the robustness of the proteomic approach in the classification of non-database bacteria. This also provides additional confidence in our findings in which identification at the species level was correct (Figure 5a-b). Based on these results, the unique sets of proteins for *Y. pestis* A1122 mostly resemble those found in the identified avirulent *Y. pestis* 91001 strain.

Figure 5b shows the identification result for the OMP extracts from both *Y. pestis* strains. This figure indicates a correct strain level identification of the studied samples. A closer look at the set of the unique protein biomarkers for virulent *Y. pestis*, shows the presence of biomarkers associated with virulence factors. For example, proteins encoded by virulence plasmids in *Y. pestis* such as pPCP1 that encodes for plasminogen activator, pCD1 that encodes for low-calcium response and pMT1 that encodes for murine toxin, the structural gene for fraction 1 protein capsule, were present. The latter protein was present in higher abundance than that of the other mentioned protein biomarkers. *Y. pestis* A1122 lacks the pCD1 plasmid and therefore did not express the corresponding OMPs encoded by the plasmid.

Comparing the number of unique proteins for the employed *Y. pestis* strains showed a difference between *Y. pestis* CO92 and *Y. pestis* A1122. The former strain had 191 unique proteins versus 89 for the latter. Upon removing the highly conserved, house-keeping and energy transfer proteins from both sets, the number of strain unique proteins for *Y. pestis* CO92 was higher than that for *Y. pestis* A1122. The protein biomarkers that were observed for virulent *Y. pestis* versus *Y. pestis* A1122 were present upon replicate analyses of the OMP extracts under different sample preparation conditions and instrumental analyses parameters.

Table 2 shows the comparison of the strain unique proteins for *Y. pestis* identified from the different cellular extracts. Also comparing whole cell lysate versus OMP extract showed a variation in the number of strain unique protein biomarkers in OMPs versus whole cell lysates. The number of strain unique proteins was slightly higher for pathogenic *Y. pestis* from OMPs extracts versus whole cell lysate. Few unique biomarkers were shared and virulence factors seemed to be present in higher abundance in the OMP extracts than in the whole cell lysates, which is in support of reported literature. (14).

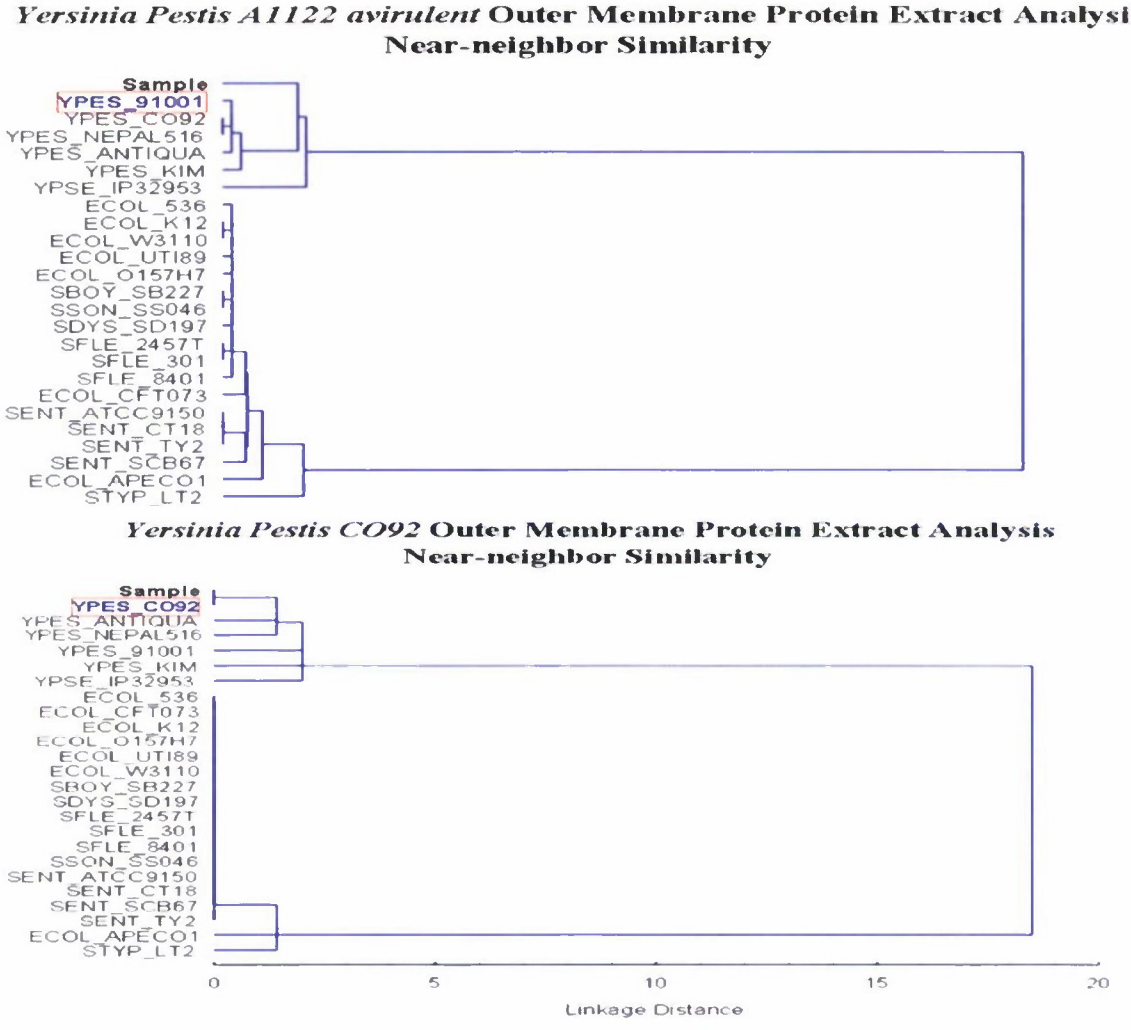


Figure 5 a-b. Near-neighbor Classification of Non-pathogenic *Y. pestis* A1122 (Figure 5a) vs. Pathogenic *Y. pestis* (Figure 5b) Using OMP Extracts.

Table 2. Unique Proteins for *Y. pestis* from Whole Cell Lysate vs. OMP Extracts

| *Y. pestis CO92* Unique proteins-Whole Cell Extract | *Y. pestis CO92* Unique proteins-OMPs Extracts |
|---|---|
| 30S ribosomal protein S6 | 50S ribosomal protein L5 |
| 50S ribosomal protein L32 | murine toxin |
| Acid-induced glycyl radical enzyme | attachment invasion locus protein |
| cationic 19 kDa outer membrane protein p | elongation factor Ts |
| cationic 19 kDa outer membrane protein pr | elongation factor G |
| chorismate mutase | putative outer membrane porin A protein |
| DNA-binding protein HU-alpha | hypothetical protein plu4065 |
| ferritin | major outer membrane lipoprotein |
| hypothetical protein YP_0808 | putative outer membrane porin A protein |
| hypothetical protein YP_1194 | putative outer membrane porin A protein |
| hypothetical protein YP_1779 | Acid-induced glycyl radical enzyme |
| hypothetical protein YP_1779 | attachment invasion locus protein |
| hypothetical protein YP_pCD78 | hypothetical protein y0163 |
| inorganic pyrophosphatase | chaperonin GroEL |
| major outer membrane lipoprotein | phosphopentomutase |
| outer membrane protein X | plasminogen activator |
| periplasmic maltose-binding protein | enolase |
| Protein | Protein |
| PTS system, mannose-specific IIAB compone | outer membrane protein X |
| putative cystine-binding periplasmic pro | putative lipoprotein |
| putative lipoprotein | elongation factor Ts |
| putative lipoprotein | 30S ribosomal protein S2 |
| putative oxidoreductase | cationic 19 kDa outer membrane protein pr |
| secreted thiol:disulfide interchange prot | outer membrane protein 1b (lb;c) |
| thiol peroxidase | attachment invasion locus protein |
| trigger factor | 50S ribosomal protein L9 |
| urease (urea amidohydrolase) beta subunit | attachment invasion locus protein |
| urease (urea amidohydrolase) beta subunit | attachment invasion locus protein |
| | hypothetical protein y2159 |
| | hypothetical protein YP_3210 |
| | malate dehydrogenase |
| | manganese superoxide dismutase |

## 3.4 Differentiation of Pathogenic vs. Non-pathogenic *B. anthracis* Strains Using Whole Cell Lysates.

Pathogenic and non-pathogenic *B. anthracis* Ames and Sterne, respectively, were analyzed by proteomic mass spectrometry for identification. Figure 6a-b shows the histogram for the sequence-to-bacterium binary matrix, with the number of unique peptides on the y-axis and bacterium proteome on the x-axis. As seen in this figure, correct identification of each strain was made, but with a higher confidence level for *B. anthracis* Ames than *B. anthracis* Sterne strain. This observation could likely be attributed to the fact that *B. anthracis* Ames has an additional plasmid lacking in *B. anthracis* Sterne strain (pX02) and therefore is not an expressed protein biomarker detectable in the MS-Proteomic analysis for that of *B. anthracis* Sterne. In Figure 6b,

the presence of *B. anthracis* Ames slightly above the threshold cutoff of 95% confidence level supports such an observation. This was not the case with *B. anthracis* Ames samples where a distinct identification unadulterated by the presence of other *B. anthracis* strains was observed. The strain unique peptides observed with both *B. anthracis* strains are an indication of the application spectrum of such an approach. Although *B. anthracis* strains do not possess OMPs, using their whole cell lysate was sufficient to reveal the discrimination power of the MS-based proteomic approach. It is evidence that using higher concentrations and optimization of the lysis protocols could enhance the MS-based proteomic analysis and provide a use for *Bacillus* differentiation at the strain level.

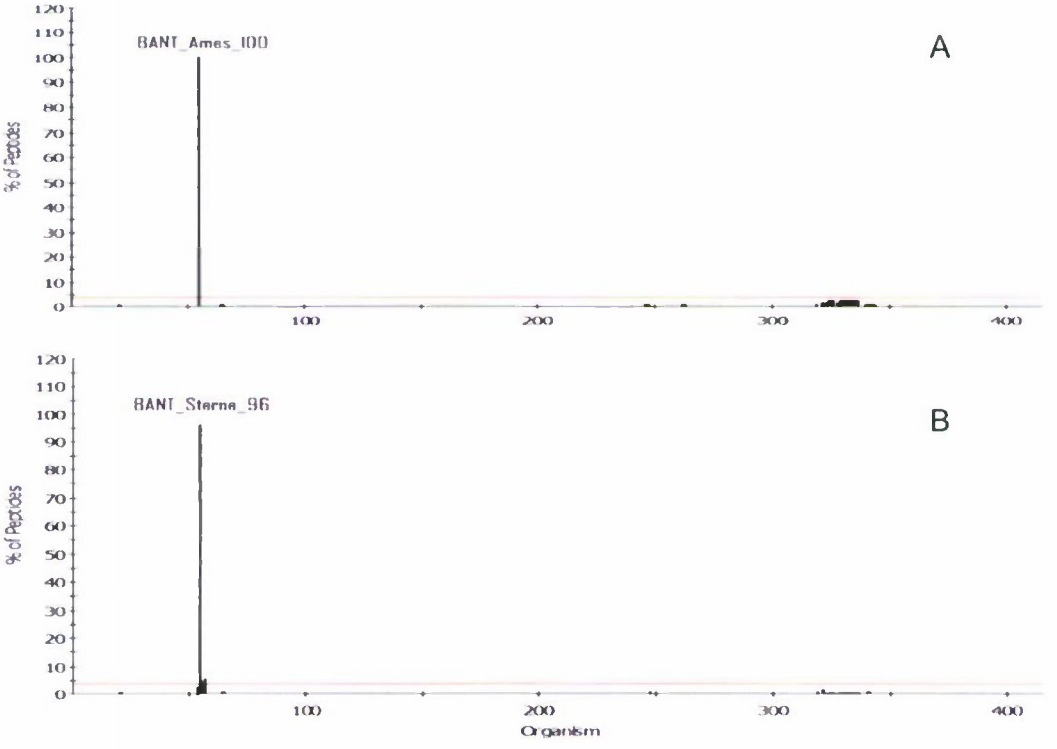


Figure 6 a-b. Bacterial Differentiation of *B. anthracis* Strains Using Whole Cell Lysates. Figure 6a represents the identification of *B. anthracis* Ames strain, while Figure 6b represents the identification of *B. anthracis* Sterne strain. X-axis represents bacterium proteome and Y-axis represents number of unique peptides at 95% confidence level. The horizontal line is the threshold under which peptides identified are considered statistically non-significant.

Further work should be conducted to investigate the strain unique peptides for *B. anthracis* strains to understand their characterization in a biologically meaningful medium.

Doing so will increase our knowledge of the set of functional proteins responsible for strain virulence. Determining their corresponding set of genes that can be biological manipulated under different environmental conditions will be of great interest to measure the validity of strain differentiation.

## 4. CONCLUSIONS

This project revealed the advantage of using OMPs as unique biomarkers for bacterial differentiation of pathogenic versus nonpathogenic strains. The differentiation capability enhanced the confidence level of the discrimination process through the utilization of OMPs as biomarkers. OMPs provide a unique source of cellular variability and thus, introduce biodiversity among cellular proteins for very similar bacterial strains and thereby provide distinct and unique protein biomarkers. The whole cell lysates did provide discrimination; however, the possible ionization suppression could shield the detection of important peptides that could be classified as unique biomarkers. On the other hand, whole cell lysates are an appropriate option for the differentiation of gram-positive bacterial strains and the reported results herein support their potential application in differentiation. Overall, an extension of this project to include a wider investigation of relevant pathogenic bacteria such as *Francisella tularensis, Burkholderia spp.* and other relevant strains could provide us with a more global outlook of the importance of OMPs with regard to pathogenicity and how we can confidently identify organisms at the strain level using protein biomarkers for bacterial classification and diagnostic purposes.

LITERATURE CITED

1. Dworzanski, J.P.; Deshpande, S.V.; Chen, R.; Jabbour, R.E.; Snyder, A.P.; Wick, C.H.; Li., L. Mass Spectrometry-Based Proteomics Combined with Bioinformatic Tools for Bacterial Classification. *J. Proteome Res.* **2006**, *5*, 76.

2. Garmendia, J.; Frankel, G.; Crepin, V.F. Enteropathogenic and Enterohemorrhagic *Escherichia coli* Infections: Translocation, Translocation, Translocation. *Infect. Immun.* **2005**, *73*, 2573-2585.

3. Ide, T.; Laarmann, S.; Greune, L.; Schillers, H.; Oberleithner, H.; Schmidt, M.A. Characterization of Translocation Pores Inserted into Plasma Membranes by Type III-Secreted Esp Proteins of Enteropathogenic *Escherichia coli*. *Cell. Microbiol.* **2001**, *3*, 669-679.

4. McDaniel, T.K.; Kaper, J.B. A Cloned Pathogenicity Island from Enteropathogenic *Escherichia coli* Confers the Attaching and Effacing Phenotype on *E. coli* K-12. *Mol. Microbiol.* **1997**, *23*:399-407.

5. McDaniel, T.K.; Jarvis, K.G.; Donnenberg, M.S.; Kaper, J.B. A Genetic Locus of Enterocyte Effacement Conserved among Diverse Enterobacterial Pathogens. *PNAS.* **1995**, *92*, 1664-1668.

6. Wachter, C.; Beinke, C.; Mattes, M.; Schmidt, M.A. Insertion of EspD into Epithelial Target Cell Membranes by Infecting Enteropathogenic *Escherichia coli*. *Mol. Microbiol.* **1999**, *31*, 1695-1702.

7. Ben-Gurion, R; Shafferman, A. Essential Virulence Determinants of Different Yersinia Species are Carried on a Common Plasmid. *Plasmid* **1981**, *5*, 183-187.

8. Ferber, D.M.; Brubaker, R.R. Plasmids in Yersinia Pestis. *Infect. Immun.* **1981**, *31*, 839-841.

9. Filippov, A.A.; Solodovnikov, N.S.; Kookleva, L.M.; Protsenko, O.A. Plasmid Content in *Yersinia Pestis* Strains of Different Origin. *FEMS Microbiol. Lett.* **1990**, *67*, 45-48.

10. Cornelis, G.R. Yersinia Type III Secretion: Send in the Effectors. *J. Cell Biol.* **2002**, *158*, 401-408.

11. Ramamurthi, K.S.; Schneewind, O. Type III Protein Secretion in Yersinia Species. *Annu. Rev. Cell Dev. Biol.* **2002**, *18*, 107-133.

12. Titball, R.W.; Hill, J.; Lawton, D.G.; Brown, K.A. *Yersinia Pestis* and Plague. *Biochem. Soc. Trans.* **2003**, *31*, 104-107.

13. Blattner, F.R.; Plunkett, G., 3rd; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; Collado-Vides, J.; Glasner, J.D.; Rode, C.K.; Mayhew, G.F.; Gregor, J.; Davis, N.W.; Kirkpatrick, H.A.; Goeden, M.A.; Rose, D.J.; Mau, B.; Shao, Y. The Complete Genome Sequence of *Escherichia coli* K-12. *Science* **1997**, *277*, 1453-1474.

14. Tsukano, H.; Itoh, K.; Suzuki, S.; Watanabe, H. Detection and Identification of *Yersinia Pestis* by Polymerase Chain Reaction (PCR) using Multiplex Primers. *Microbiol. Immunol.* **1996**, *40*, 773-775.